

# Bounding the Coarse Graining Error in Hidden Markov Dynamics

David Andrieux

*Center for Nonlinear Phenomena and Complex Systems,  
Université Libre de Bruxelles, B-1050 Brussels, Belgium*

Lumping a Markov process introduces a coarser level of description that is useful in many contexts and applications. The dynamics on the coarse grained states is often approximated by its Markovian component. In this letter we derive finite-time bounds on the error in this approximation. These results hold for non-reversible dynamics and for probabilistic mappings between microscopic and coarse grained states.

## I. INTRODUCTION

Markov processes are a standard modeling tool used in many applications ranging from finance [1] and telecommunications [2] to physics [3], chemistry [4], biology [5], and computer science [6]. For theoretical and practical reasons, it is often convenient to partition the state space into aggregates and to view the dynamics at a coarser level. The coarse graining operation bridges the gap between different level of descriptions by introducing “mesostates” each representing many microstates. Analysis of experimental data naturally leads to coarse graining as observation techniques may not be able to resolve the set of microscopic states and only give access to mesostates. Similarly, disregarding the environment or part of a system provides an effective description of a remaining (sub)system of interest.

A deterministic mapping between micro and mesostates is too restrictive to be applicable in many problems of interest. The concept of coarse graining can be extended to include the case where the mapping is a probabilistic function of the microstates. The resulting model, called a hidden Markov model, is a doubly embedded stochastic process [7]. Such processes are especially known for their application in temporal pattern recognition such as speech and handwriting recognition [8] or bioinformatics [9].

Although the dynamics on the aggregates is not Markovian in general [10], there is a natural choice for a Markov dynamics on the set of lumped states [10–12]. This dynamics reproduces the influence of the first past state on the transition probabilities and neglects the higher-order memory effects. This choice matches the time evolution of the original unlumped state started at the stationary state. Furthermore, the probability transfers between aggregates match those arising from the original chain.

In the present work we analyze the accuracy of such coarse grained models as compared to the exact microscopic behavior. This problem was first envisaged by Hoffman and Salamon in [12] for the special case of deterministic coarse gaining and reversible Markov chains. Reversible Markov chains have transition matrices diagonally similar to symmetric matrices, a strong symmetry property at the basis of their analysis. Here we generalize their approach to the case of probabilistic coarse graining and non-reversible dynamics. We obtain bounds on the error made by using the lumped dynamics considered as a model of the unlumped dynamics.

## II. LUMPED MARKOV CHAINS

We consider a Markov chain characterized by a transition matrix  $G$  on the finite state space  $\Sigma$ . The probability distribution  $\mathbf{p} = (p_1, p_2, \dots, p_N)$  evolves in discrete time steps according to

$$\mathbf{p}(n+1) = \mathbf{p}(n)G. \quad (1)$$

We assume that the Markov chain is primitive, i.e., there exists an  $n_0$  such that  $G^{n_0}$  has all positive entries. This guarantees that  $G$  has a unique stationary distribution  $\boldsymbol{\pi}$  such that

$$\boldsymbol{\pi} = \boldsymbol{\pi}G. \quad (2)$$

Our goal is to analyze lumped dynamics. Let  $\{\omega_j\} \in \Omega$  be the set of mesostates. The mesostate  $\omega$  is observed with probability  $b_i(\omega)$  when the system is in microstate  $i$ . We collect these conditional probability distributions into the matrix  $C$  with elements

$$C_{i\omega} = b_i(\omega). \quad (3)$$

The matrix  $C$  serves to specify the lumped probability distribution  $\hat{\mathbf{p}} = \mathbf{p}C$  on  $\Omega$  corresponding to a distribution  $\mathbf{p}$  on  $\Sigma$ . We also introduce the matrix  $D$  with elements

$$D_{\omega i} = \frac{\pi_i b_i(\omega)}{\sum_k \pi_k b_k(\omega)}. \quad (4)$$

The element  $D_{\omega i}$  is the conditional probability to be in state  $i$  given the observation  $\omega$ . In the case of a deterministic association between microstates and aggregates, the operators  $C$  and  $D$  reduce to the operators introduced in [12]. Their successive action defines a stochastic operator  $CD$  that satisfies

$$\pi = \pi CD. \quad (5)$$

Following [11], we now introduce the lumped dynamics with transition matrix

$$\hat{G} = DGC. \quad (6)$$

This matrix is stochastic,  $\hat{G} \geq 0$  and  $\sum_{\omega'} \hat{G}_{\omega\omega'} = 1$ . This choice of the transition matrix insures that the distribution  $\hat{\pi} = \pi C$  is the stationary distribution of  $\hat{G}$ :

$$\hat{\pi} \hat{G} = \pi CDGC = \pi GC = \pi C = \hat{\pi}. \quad (7)$$

By construction, the dynamics  $\hat{G}$  also preserves the probability fluxes between states in the coarse grained description. Precisely, the dynamics  $\hat{G}$  arises from the Markovian approximation of a stationary sequence of observed mesostates.  $\hat{G}$  is the unique Markov chain on  $\Omega$  that satisfies this condition.

### III. BOUNDING COARSE GRAINING ERRORS

Starting from a distribution  $\mathbf{p}_0$  on  $\Sigma$ , its time evolution among the aggregates with the unlumped dynamics is  $\mathbf{p}_0 G^n C$ , while its time evolution with the lumped dynamics is  $\mathbf{p}_0 C \hat{G}^n$ . The main question considered here is how different these two dynamics can be. To address this question, we define the norm  $\|\mathbf{v}\|_\pi = \|\mathbf{v} U_\pi^{-1}\|_2$ , where  $U_\pi = \text{diag}(\sqrt{\pi_1}, \sqrt{\pi_2}, \dots, \sqrt{\pi_N})$  and  $\|\cdot\|_2$  is the 2-norm. The corresponding operator norm is

$$\|A\|_\pi = \|U_\pi A U_\pi^{-1}\|_2. \quad (8)$$

The difference between the two probability distributions after  $n$  time steps can be expressed as

$$\begin{aligned} \left\| \mathbf{p}_0 C \hat{G}^n - \mathbf{p}_0 G^n C \right\|_\pi &= \left\| \mathbf{p}_0 C (DGC)^n - \mathbf{p}_0 G^n C \right\|_\pi \\ &= \left\| \mathbf{p}_0 (CDG)^n C - \mathbf{p}_0 G^n C \right\|_\pi \\ &= \left\| \mathbf{p}_0 ((CDG)^n - G^n) C \right\|_\pi. \end{aligned} \quad (9)$$

As emphasized in [12], we observe the prominent role of the operator  $CDG \equiv H$ , which specifies a dynamics on the original state space  $\Sigma$ .

Our goal will be to bound the  $n$ -step difference  $\|H^n - G^n\|_\pi$ . The  $n$ -step difference can transiently grow, but must eventually decline to zero as, by construction, the lumped and the unlumped chain converge to the same stationary distribution.

We use of the fact that  $H$  and  $G$  have the common stationary distribution  $\pi$ . We define the projection operator  $P_\pi = \mathbf{u}^T \pi$ , where  $\mathbf{u}$  is the vector  $(1, 1, \dots, 1) \in \mathbb{R}^N$ . The complementary projection  $P_\sigma = I - P_\pi$ . We end up with the following representation of  $G$  and  $H$ :

$$G = (P_\pi + P_\sigma)G = P_\pi + P_\sigma G \quad (10)$$

and

$$H = (P_\pi + P_\sigma)H = P_\pi + P_\sigma H. \quad (11)$$

From (10) and (11) we obtain  $H - G = P_\sigma H - P_\sigma G$ , and  $P_\pi P_\sigma G = P_\sigma G P_\pi = P_\pi P_\sigma H = P_\sigma H P_\pi = 0$ .

The norms  $\|P_\sigma H\|_\pi$  and  $\|P_\sigma G\|_\pi$  will play a crucial role in our analysis. In this regard, we prove the following theorem.

**Theorem 1.** *Let  $G$  be a transition probability matrix, and let  $P_\sigma$  the projection operator introduced above. Then*

$$\|P_\sigma G\|_\pi < 1. \quad (12)$$

Furthermore,  $\|P_\sigma G\|_\pi = \sigma_2$ , where  $\sigma_2$  is the second-largest singular value of  $U_\pi G U_\pi^{-1}$ .

PROOF OF THEOREM 1. We will use the following result [13]. Let  $A$  be a matrix with nonnegative entries and spectral radius  $\rho(A)$ , and suppose that there exist left and right positive Perron eigenvectors  $\mathbf{v}$  and  $\mathbf{w}$ , respectively. Then  $\rho(A) = \|XAX^{-1}\|_2$ , where  $X = \text{diag}(v_j^{1/2}w_j^{-1/2})$ .

In our case this translates into  $\|G\|_\pi = \|U_\pi G U_\pi^{-1}\|_2 = \rho(G) = 1$ . Because the 2-norm of a matrix  $A$  is given by its dominant singular value, we deduce that the dominant singular value  $\sigma_1$  of  $U_\pi G U_\pi^{-1}$  equals 1. Equivalently, we have that the dominant eigenvalue of  $(U_\pi G U_\pi^{-1})(U_\pi G U_\pi^{-1})^T$  is  $\sigma_1 = 1$ .

We now turn to the norm  $\|P_\sigma G\|_\pi$ . Note that  $P_\sigma G \mathbf{u}^T = (G - P_\pi) \mathbf{u}^T = 0$ , so that  $P_\sigma G$  has negative elements and the above construction cannot be applied. Introducing the notation  $U_\pi A U_\pi^{-1} \equiv \bar{A}$ , the norm  $\|P_\sigma G\|_\pi = \|\bar{P}_\sigma \bar{G}\|_2$ . It is thus given by the largest eigenvalue of  $\bar{P}_\sigma \bar{G} \bar{G}^T \bar{P}_\sigma^T$ .

First, we consider the projection operator  $\bar{P}_\pi$ . A direct calculation shows that  $P_\pi$  is symmetrized by  $U_\pi^{-1}$ . It follows that  $\bar{P}_\sigma = I - \bar{P}_\pi$  is symmetric as well. Now, because  $G$  and  $P_\sigma$  commute,  $\bar{G}$  and  $\bar{P}_\sigma$  commute with each other. Accordingly,  $\bar{P}_\sigma \bar{G} \bar{G}^T \bar{P}_\sigma^T = \bar{P}_\sigma \bar{G} \bar{G}^T$ . Furthermore, there exists an eigenbasis such that the eigenvalues of  $\bar{P}_\sigma \bar{G} \bar{G}^T$  take the form  $\alpha_i \beta_i$ , where  $\alpha_i$  and  $\beta_i$  are the eigenvalues of  $\bar{P}_\sigma$  and  $\bar{G} \bar{G}^T$ , respectively. Because  $\bar{P}_\sigma$  is a projection operator, it has  $(N-1)$  eigenvalues 1 and one eigenvalue 0. The latter corresponds to the right eigenvector  $\sqrt{\pi_i}$ .

The vector  $\sqrt{\pi_i}$  is a right eigenvector of  $\bar{G} \bar{G}^T$  with eigenvalue 1. From  $\sigma_1 = 1$ , we have that 1 is the dominant eigenvalue of  $\bar{G} \bar{G}^T$ . Therefore, the eigenvalues of  $\bar{P}_\sigma \bar{G} \bar{G}^T$  are given by the eigenvalues of  $\bar{G} \bar{G}^T$ , except for its dominant eigenvalue 1 that is replaced by 0. In particular,  $\|P_\sigma G\|_\pi$  is given by the dominant eigenvalue of  $\bar{P}_\sigma \bar{G} \bar{G}^T$  or, equivalently, by the second-largest singular value  $\sigma_2$  of  $\bar{G} = U_\pi G U_\pi^{-1}$ .

We now have to prove that  $\sigma_2 < 1$ . We already know that  $\sigma_2 \leq \sigma_1 = 1$ , but we now show that the inequality is strict. This follows from the Perron-Frobenius theorem applied to  $\bar{G} \bar{G}^T$ . Indeed, because  $G$  is a primitive transition matrix,  $\bar{G} \bar{G}^T$  is nonnegative and primitive. Recalling that the norm of an operator is always greater or equal to its spectral radius,  $\rho(A) \leq \|A\|$ , we arrive at

$$|\lambda_2| \leq \|P_\sigma G\|_\pi = \sigma_2 < 1, \quad (13)$$

where  $\lambda_2$  is the second-largest (in modulus) eigenvalue of  $G$ .  $\square$

Because  $H$  is a transition matrix, we deduce from Theorem 1 that  $\|P_\sigma H\|_\pi = \eta_2 < 1$ , with  $\eta_2$  the second-largest singular value of  $U_\pi H U_\pi^{-1}$ . Furthermore, we have

$$\begin{aligned} \|P_\sigma H\|_\pi &= \|CDG - P_\pi\|_\pi = \|CDG - CDP_\pi\|_\pi \\ &= \|CD(G - P_\pi)\|_\pi \leq \|CD\|_\pi \|P_\sigma G\|_\pi. \end{aligned} \quad (14)$$

Noting that  $CD$  is stochastic and that the similarity transform  $U_\pi^{-1}$  symmetrizes  $CD$ , we conclude that  $\|CD\|_\pi = 1$ . This leads to

$$\|P_\sigma H\|_\pi \leq \|P_\sigma G\|_\pi \quad \text{or} \quad \eta_2 \leq \sigma_2. \quad (15)$$

We are now in position to derive our first bound. The following theorem bounds the  $n$ -step difference  $\|H^n - G^n\|_\pi$  in terms of the one-step difference  $\|H - G\|_\pi$ .

**Theorem 2.** *Let  $G$  be a transition probability matrix, and let  $H = CDG$  with  $C$  and  $D$  as introduced above. Define*

$$\delta = \|H - G\|_\pi. \quad (16)$$

*Then*

$$\|H^n - G^n\|_\pi \leq \delta K'(n) \leq \delta K(n), \quad (17)$$

*where  $K'(n) = (\sigma_2^n - \eta_2^n)/(\sigma_2 - \eta_2)$  and  $K(n) = n\sigma_2^{n-1}$ . Here  $\sigma_2$  and  $\eta_2$  are the second-largest singular values of  $U_\pi G U_\pi^{-1}$  and  $U_\pi H U_\pi^{-1}$ , respectively.*

PROOF OF THEOREM 2. We start by expressing the  $n$ -step difference in terms of the one-step difference as in [12]. We note that

$$\|H^n - G^n\|_\pi = \|(H - G)H^{n-1} + G(H^{n-1} - G^{n-1})\|_\pi. \quad (18)$$

Iterating, we find

$$\begin{aligned} \|H^n - G^n\|_\pi &= \left\| \sum_{k=0}^{n-1} G^k (H - G) H^{n-k-1} \right\|_\pi \\ &\leq \sum_{k=0}^{n-1} \|G^k (H - G) H^{n-k-1}\|_\pi. \end{aligned} \quad (19)$$

Then for integers  $k, n$ ,

$$\begin{aligned} \|G^k(H - G)H^{n-k-1}\|_\pi &= \|(P_\pi + P_\sigma G)^k(P_\sigma H - P_\sigma G)(P_\pi + P_\sigma H)^{n-k-1}\|_\pi \\ &= \|(P_\sigma G)^k(P_\sigma H - P_\sigma G)(P_\sigma H)^{n-k-1}\|_\pi \\ &\leq \|P_\sigma G\|_\pi^k \delta \|P_\sigma H\|_\pi^{n-k-1}. \end{aligned} \quad (20)$$

Combining (19) and (20) we find

$$\|H^n - G^n\|_\pi \leq \delta \sum_{k=0}^{n-1} \|P_\sigma G\|_\pi^k \|P_\sigma H\|_\pi^{n-k-1}. \quad (21)$$

Carrying out the summation in (21) and using Theorem 1 we obtain the bound

$$\|H^n - G^n\|_\pi \leq \delta K'(n) = \delta \frac{\sigma_2^n - \eta_2^n}{\sigma_2 - \eta_2}. \quad (22)$$

This expression can be bound from above by combining (21) and (14) to give

$$K'(n) \leq K(n) = n\sigma_2^{n-1}, \quad (23)$$

with equality when  $\eta_2 = \sigma_2$ .  $\square$

We now derive a bound independent of the one-step difference  $\|H - G\|_\pi$ .

**Theorem 3.** *Let  $G$  be a transition probability matrix, and let  $H = CDG$  with  $C$  and  $D$  as introduced above. Then*

$$\|H^n - G^n\|_\pi \leq \eta_2^n + \sigma_2^n \leq 2\sigma_2^n, \quad (24)$$

where  $\sigma_2$  and  $\eta_2$  are the second-largest singular values of  $U_\pi G U_\pi^{-1}$  and  $U_\pi H U_\pi^{-1}$ , respectively.

**PROOF OF THEOREM 3.** We observe that  $G^n = (P_\pi + P_\sigma G)^n = P_\pi + (P_\sigma G)^n$  and  $H^n = (P_\pi + P_\sigma H)^n = P_\pi + (P_\sigma H)^n$ . Hence we have

$$\begin{aligned} \|H^n - G^n\|_\pi &= \|(P_\sigma H)^n - (P_\sigma G)^n\|_\pi \\ &\leq \|(P_\sigma H)^n\|_\pi + \|(P_\sigma G)^n\|_\pi \\ &\leq \|P_\sigma H\|_\pi^n + \|P_\sigma G\|_\pi^n. \end{aligned} \quad (25)$$

From Theorem 1 and  $\|P_\sigma H\|_\pi \leq \|P_\sigma G\|_\pi$  we deduce the inequalities (24).  $\square$

We end this section with the following remarks. The bound  $2\sigma_2^n$  from Theorem 3 is independent of  $\delta$  and  $\eta_2$ . Accordingly, it is valid for any probabilistic coarse graining of the original chain. This bound also shows that the coarse graining error decreases exponentially in time. Relation (13) reveals that the fastest possible decay rate of the bound is  $|\lambda_2|$ . We show in the next section that this rate is achieved for special classes of Markov dynamics.

## IV. SPECIAL MARKOV CHAINS

### A. Reversible Markov chains

A Markov chain is reversible if it obeys the detailed balance conditions

$$\pi_i G_{ij} = \pi_j G_{ji} \quad \forall i, j \in \Sigma. \quad (26)$$

This corresponds to an equilibrium situation where no probability currents are present in the stationary state. We note that in this case  $\hat{G}$  is also reversible. Under the detailed balance conditions (26) the operator  $U_\pi P_\sigma G U_\pi^{-1}$  is symmetric. This strong symmetry property of reversible Markov chains is at the basis of Hoffman and Salamon's analysis. Accordingly we have

$$\|P_\sigma G\|_\pi = |\lambda_2|, \quad (27)$$

where  $\lambda_2$  is the second-largest eigenvalue of  $G$ . In this way we recover the bound  $K(n) = n|\lambda_2|^{n-1}$  of [12].

## B. Doubly stochastic matrices

Doubly stochastic matrices are characterized by the following property:

$$\sum_i G_{ij} = 1 \quad \forall j \in \Sigma, \quad (28)$$

i.e., both their rows and columns sum to one. This implies that the stationary distribution is uniform,  $\boldsymbol{\pi} = (1/N, \dots, 1/N)$ . In particular, we have that  $A = U_\pi A U_\pi^{-1}$  for any operator  $A$ . Doubly stochastic matrices do not necessarily satisfy the reversibility conditions (26). Notably, the class of doubly stochastic matrices coincides with the class of normal stochastic matrices [14]. Normal matrices commute with their transpose and have their eigenvalues as singular values. Noting that  $P_\sigma G$  is normal if  $G$  is, we conclude that  $\|P_\sigma G\|_\pi = \|P_\sigma G\|_2 = \sigma_2 = |\lambda_2|$ , yielding

$$K(n) = n|\lambda_2|^{n-1}. \quad (29)$$

## V. CONTINUOUS-TIME MARKOV PROCESSES

The previous construction can be extended to continuous-time Markov processes using the concept of uniformization [15, 16].

The probability distribution  $\boldsymbol{p}(t)$  now obeys the dynamics

$$\frac{d\boldsymbol{p}(t)}{dt} = \boldsymbol{p}(t)L, \quad (30)$$

with the rate matrix  $L_{ij} \geq 0$  for  $i \neq j$ , and  $L_{ii} = -\sum_{j \neq i} L_{ij}$ . Note that the rate matrix has negative elements. We assume it has a unique stationary distribution  $\boldsymbol{\pi}$  such that  $0 = \boldsymbol{\pi}L$ .

We define the matrices  $C$  and  $D$  as above. The lumped dynamics  $\hat{L} = DLC$  is verified to be a rate matrix:  $\hat{L}_{\omega\omega'} \geq 0$  for  $\omega \neq \omega'$  and  $\hat{L}_{\omega\omega} = -\sum_{\omega' \neq \omega} \hat{L}_{\omega\omega'}$ .

Starting from a distribution  $\boldsymbol{p}_0$ , the difference between the two dynamics after a time  $t$  reads

$$\left\| \boldsymbol{p}_0 C e^{t\hat{L}} - \boldsymbol{p}_0 e^{tL} C \right\|_\pi = \left\| \boldsymbol{p}_0 (e^{tH} - e^{tL}) C \right\|_\pi, \quad (31)$$

where we defined the operator  $CDL \equiv H$ .

We now introduce the transition matrix  $T(\beta) = I + L/\beta$ , where  $I$  is the unity operator and  $\beta$  the uniformization parameter [15, 16]. To ensure that  $T(\beta)$  is a proper transition matrix,  $\beta$  must satisfy  $\beta \geq \max_i |L_{ii}|$ . The distribution  $\boldsymbol{\pi}$  is also the stationary distribution of  $T(\beta)$ ,  $\boldsymbol{\pi}T(\beta) = \boldsymbol{\pi}(I + L/\beta) = \boldsymbol{\pi}$ , for all  $\beta$ .

We thus have

$$\begin{aligned} \|e^{tL}\|_\pi &= \left\| e^{t\beta(T-I)} \right\|_\pi \\ &= \left\| e^{t\beta P_\sigma(T-I)} \right\|_\pi \\ &= \left\| e^{-t\beta P_\sigma} e^{t\beta P_\sigma T} \right\|_\pi \\ &\leq \left\| e^{-t\beta P_\sigma} \right\|_\pi \left\| e^{t\beta P_\sigma T} \right\|_\pi \\ &\leq e^{-t\beta \|P_\sigma\|_\pi} e^{t\beta \|P_\sigma T\|_\pi} \\ &= e^{-t\beta[1-\sigma_2(\beta)]}. \end{aligned} \quad (32)$$

In the second line we used that  $P_\sigma(T - I) = T - I$ , in the third line that  $P_\sigma$  commutes with  $P_\sigma T$ , and in the last equality that  $\|P_\sigma\|_\pi = 1$ . Here  $\sigma_2(\beta) < 1$  is the second-largest singular value of  $U_\pi T(\beta) U_\pi^{-1}$ .

We have  $\|H\|_\pi \leq \|CD\|_\pi \|L\|_\pi = \|L\|_\pi$ , from which we deduce  $\exp(t\|H\|_\pi) \leq \exp(t\|L\|_\pi)$ . Taking into account (32) we obtain the bound

$$\left\| e^{tH} - e^{tL} \right\|_\pi \leq 2e^{-\beta t[1-\sigma_2(\beta)]}, \quad (33)$$

which decreases exponentially in time. This bound can be further optimized by minimizing over the uniformization parameter  $\beta$ .

## VI. CONCLUSIONS

We derived quantitative bounds on the error made by using a lumped Markov process instead of the unlumped dynamics. Notably, the deviations between the two levels of description can be uniformly bounded in terms of their deviation in one time step. The bounds are expressed in terms of the second-largest singular values of the transition probability matrices. These results generalize the work by Hoffman and Salamon [12] for reversible Markov chains and deterministic coarse graining. Our construction holds for discrete- and continuous-time, and for non-reversible processes and probabilistic coarse graining. The important finding is that our bounds hold for all time and are not just asymptotic.

The main technique making our bounds possible consisted in the use of a carefully chosen operator norm. Exploiting the fact that transition matrices are nonnegative, we find a norm that equals the dominant singular value. On the other hand, important observables such as the statistics of current fluctuations are described in terms of generalized transition operators that are nonnegative but non-stochastic [16]. As our approach relies on the fact that the singular values of stochastic matrices are lower than one, it is not clear how to extend our arguments to these non-stochastic operators. In addition, the impact of coarse graining on observables nonlinear in the probability distribution such as the entropy production remains to be investigated [17].

Other dynamics on the aggregates that are consistent with the stationary state  $\hat{\pi}$  can be defined. These dynamics can satisfy further requirements, such as that the *net* probability transfers between aggregates match those derived from the unlumped chain. The “gauge” freedom available in choosing the dynamics might be used to minimize the coarse graining error while preserving relevant dynamical features.

**Acknowledgments.** This work is supported by the F.R.S.-FNRS Belgium.

- 
- [1] M. Kijima, *Stochastic Processes with Applications to Finance* (Chapman and Hall, 2002).
  - [2] G. Giambene, *Queuing Theory and Telecommunications: Networks and Applications* (Springer, 2010).
  - [3] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
  - [4] C. Gardiner, *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences* (Springer, 2004).
  - [5] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology* (Chapman and Hall, 2010).
  - [6] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications* (Wiley-Interscience, 2006).
  - [7] L. E. Baum and T. Petrie, *Ann. Math. Stat.* **37**, 1554 (1966).
  - [8] G. A. Fink, *Markov Models for Pattern Recognition: From Theory to Applications* (Springer, 2010).
  - [9] T. Koski, *Hidden Markov Models of Bioinformatics* (Springer, 2002).
  - [10] J. G. Kemeny and J. L. Snell (Eds.), *Finite Markov chains* (Van Nostrand, Princeton, NJ, 1960).
  - [11] C. J. Burke and M. Rosenblatt, *Ann. Math. Statist.* **29**, 1112 (1958).
  - [12] K. H. Hoffmann and P. Salamon, *Appl. Math. Lett.* **22**, 1471 (2009).
  - [13] D. Hershkowitz, W. Huang, H. Schneider, and H. Weinberger, *SIAM J. Mat. Anal. Appl.* **18**, 249 (1997).
  - [14] R. Sinkhorn, *Linear Algebra Appl.* **40**, 225 (1981).
  - [15] A. Jensen, *Skand. Aktuarretidskr* **36**, 87 (1953).
  - [16] D. Andrieux, *Phys. Rev. E* **82**, 031124 (2010).
  - [17] G. Nocolis, *Phys. Rev. E* **83**, 011112 (2011).